designer amino acids (*e.g.*, β-methyl amino acids, α-methyl amino acids, Nα-methyl amino acids) and amino acid analogs in general. In addition, when an α-carbon atom has four different groups (as is the case with the 20 amino acids used by biological systems to synthesize proteins, except for glycine, which has two hydrogen atoms bonded to the α carbon atom), two different enantiomeric forms of each amino acid exist, designated D and L. In mammals, only L-amino acids are incorporated into naturally occurring polypeptides. Of course, the instant invention envisions proteins incorporating one or more D- and L- amino acids, as well as proteins comprised of just D- or L- amino acid residues.

As used herein, a "β-carbon atom" refers to the carbon atom (if present) in the R group of the side chain of an amino acid (or amino acid residue) that is covalently bonded to the α-carbon atom of that amino acid (or residue). For purposes of this invention, glycine is the only naturally occurring amino acid found in mammalian proteins that does not contain a β-carbon atom.

A "side chain center of mass" of an amino acid or amino acid residue refers to the calculated position in three-dimensional space of the center of mass of the sum total of the masses of all atoms comprising that side chain, although it may also include the alpha carbon and/or amino nitrogen of a particular amino acid or residue thereof. Herein, a side chain center of mass is preferably represented as a single pseudoatom.

Conventional amino acid residue abbreviations are used throughout this patent, and both the one and three letter codes are reproduced here for convenience: alanine = "A" or "Ala"; arginine = "R" or "Arg"; asparagine = "N" or "Asn"; aspartic acid = "D" or "Asp"; cysteine = "C" or "Cys"; glutamic acid = "E" or "Glu" glutamine = "Q" or "Gln"; glycine = "G" or "Gly"; histidine = "H" or "His"; isoleucine = "I" or "Ile"; leucine = "L" or "Leu"; lysine "K" or "Lys"; methionine = "M" or "Met"; phenylalanine = "F" of "Phe"; proline "P" or "Pro"; serine = "S" or "Ser"; threonine = "T" or "Thr"; tryptophan = "W" or "Trp"; tyrosine = "Y" or "Tyr"; and valine = "V" or "Val". Amino acid sequences are written from carboxy-

to amino-terminus, unless otherwise indicated. Conventional nucleic acid nomenclature is also used, wherein "A" means adenine, "C" means cytosine, "G" means guanine, "T" means thymine, and "U" means uracil. Nucleotide sequences are written from 5' to 3', unless otherwise indicated.

"Protein" refers to any polymer of two or more individual amino acids (whether or not naturally occurring) linked via a peptide bond, and occurs when the carboxyl carbon atom of the carboxylic acid group bonded to the α-carbon of one amino acid (or amino acid residue) becomes covalently bound to the amino nitrogen atom of amino group bonded to the α-carbon of an adjacent amino acid. These peptide bond linkages, and the atoms comprising them (*i.e.*, α-carbon atoms, carboxyl carbon atoms (and their substituent oxygen atoms), and amino nitrogen atoms (and their substituent hydrogen atoms)) form the "polypeptide backbone" of the protein. In simplest terms, the polypeptide backbone shall be understood to refer the amino nitrogen atoms, α-carbon atoms, and carboxyl carbon atoms of the protein, although two or more of these atoms (with or without their substituent atoms) may also be represented as a pseudoatom.

The term "protein" is understood to include the terms "polypeptide" and "peptide" (which, at times, may be used interchangeably herein) within its meaning. In addition, proteins comprising multiple polypeptide subunits (*e.g.*, DNA polymerase III, RNA polymerase II), as well as other non-proteinaceuos catalytic molecules (*e.g.*, ribozymes) will also be understood to be included within the meaning of "protein" as used herein. Similarly, "protein fragments," *i.e.*, stretches of amino acid residues that comprise fewer than all of the amino acid residues of a protein, are also within the scope of the invention and may be referred to herein as "proteins." Additionally, "protein domains" are also included within the term "protein." A "protein domain" represents a portion of a protein comprised of its own semi-independent folded region having its own characteristic spherical geometry with hydrophobic core and polar exterior.

In biological systems (be they *in vivo* or *in vitro*, including cell-free, systems), the particular amino acid sequence of a given protein (*i.e.*, the polypeptide's "primary structure," when written from the amino-terminus to carboxy-terminus) is determined by the nucleotide sequence of the coding portion of a messenger RNA ("mRNA") molecule, which is in turn specified by genetic information, typically plasmid or genomic DNA (which, for purposes of this invention, is understood to include organelle DNA, for example, mitochondrial DNA and chloroplast DNA, as well as forms of viral genomes integrated into the genomic DNA of a host cell). Of course, any type of nucleic acid which constitutes the genome of a particular organism (*e.g.*, double-stranded DNA in the case of most animals and plants, single or double-stranded RNA in the case of some viruses, *etc.*) is understood to code for the gene product(s) of the particular organism. Messenger RNA is translated on a ribosome, which catalyzes the polymerization of a free amino acid, the particular identity of which is specified by the particular codon (with respect to mRNA, three adjacent A, G, C, or U ribonucleotides in the mRNA's coding region) of the mRNA then being translated, to a nascent polypeptide. Recombinant DNA techniques have enabled the large-scale synthesis of polypeptides (*e.g.*, human insulin, human growth hormone, erythropoietin, granulocyte colony stimulating factor, *etc.*) having the same primary sequence as when produced naturally in living organisms. In addition, such technology has allowed the synthesis of analogs of these and other proteins, which analogs may contain one or more amino acid deletions, insertions, and/or substitutions as compared to the native proteins. Recombinant DNA technology also enables the synthesis of entirely novel proteins.

In non-biological systems (*e.g.*, those employing solid state synthesis), the primary structure of a protein (which also includes disulfide (cystine) bond locations) can be determined by the user. As a result, polypeptides having a primary structure that duplicates that of a biologically produced protein can be achieved, as